Accepted Manuscript

Modelling structural changes in the volatility process

Bart Frijns, Thorsten Lehnert, Remco C.J. Zwinkels

PII: S0927-5398(11)00006-5
DOI: doi: 10.1016/j.jempfin.2011.01.005
Reference: EMPFIN 540

To appear in: *Journal of Empirical Finance*

Received date: 12 October 2010
Revised date: 7 December 2010
Accepted date: 23 January 2011

Please cite this article as: Frijns, Bart, Lehnert, Thorsten, Zwinkels, Remco C.J., Modelling structural changes in the volatility process, *Journal of Empirical Finance* (2011), doi: 10.1016/j.jempfin.2011.01.005

Modelling structural changes in the volatility process*

Bart Frijns

Auckland University of Technology

Thorsten Lehnert

Luxemburg School of Finance

Remco C.J. Zwinkels†

Erasmus School of Economics

†Corresponding author at: Erasmus School of Economics, P.O. Box 1738, 3000DR Rotterdam, The Netherlands.E: Zwinkels@ese.eur.nl, T: +31 10 408 1428; F: +31 10 408 9165

*Part of this paper was written while Remco Zwinkels stayed at the University of Technology in Sydney, Australia. We would like to thank Tim Bollerslev, Carl Chiarella, Dick van Dijk, Tony He, participants of the 2010 European Financial Management Association Conference, 2009 Australasian Finance and Banking Conference, and the Sydney Agents seminar are acknowledged.

Abstract

GARCH-type models have been very successful in describing the volatility dynamics of financial return series for short periods of time. However, time-varying behaviour of investors, for example, may cause the structure of volatility to change and the assumption of stationarity is no longer plausible. To deal with this issue, the current paper proposes a conditional volatility model with time-varying coefficients based on a multinomial switching mechanism. By giving more weight to either the persistence or shock term in a GARCH model, conditional on their relative ability to forecast a benchmark volatility measure, the switching reinforces the persistent nature of the GARCH model. Estimation of this benchmark volatility targeting or BVT-GARCH model for Dow 30 stocks indicates that the switching model is able to outperform a number of relevant GARCH setups, both in- and out-of-sample, also without any informational advantages.

Key Words: GARCH, time varying coefficients, multinomial logit

JEL Codes: C22, G17.

1. Introduction

The GARCH methodology, by Engle (1982) and Bollerslev (1986), has been pivotal in modelling volatility ever since its introduction. It has proven its value in a number of fields, such as Value-at-Risk determination, option pricing, and, perhaps most importantly, forecasting volatility. However, one of the points of critique to the original model is its impossibility to take structural changes into account (see Diebold, 1986; Lamoureux and Lastrapes, 1990). One might argue that the time-varying nature of investor behaviour may cause the structure of volatility to change and the assumption of stationarity is no longer plausible. The observed high persistence in the volatility process may be caused by unaccounted structural breaks. Therefore, the apparent misspecification of the model is induced by a lack of flexibility. In response to this critique, several authors have proposed to introduce time-variation in the GARCH coefficients. For example, Dahlhaus and Subba Rao (2006) propose a model with time-varying parameters that is globally non-stationary, but asymptotically locally stationary. The time variation is typically modelled in terms of smooth transition models (see e.g. Amado and Teräsvirta, 2008[1]), regime switching models (see Schwert, 1989; Hamilton and Susmel, 1994; Cai, 1994; Gray, 1996), or the Spline-GARCH (see Engle and Rangel, 2008).

Such models typically perform better than standard GARCH models, both in-sample (see e.g. Gray, 1996, for an application to short-term interest rates) and out-of-sample (see e.g. Klaassen, 2002, for an application to exchange rates). However, regime switching GARCH models often require a considerable number of extra parameters and are often difficult to trace computationally. In addition, the method is purely descriptive in the sense that the regime switching probabilities do not have any intuitive foundation. In contrast, we propose a parsimonious model that still allows for time variation in the coefficients of the persistence and shock terms. By imposing a certain structure to the time variation, the results become intuitively interpretable.

In this paper we introduce time variation in the coefficients of a standard GARCH (1, 1) model using a multinominal switching mechanism, as often used in the heterogeneous agents literature (see Brock and Hommes, 1997, 1998), which is

[1]Interestingly, their findings suggest that the time-variation in the unconditional volatility induces the long-memory type behaviour.

essentially a generalization of the smooth transition autoregressive models (see Van Dijk et al., 2002, for an overview). This switching mechanism assigns weights on a one-on-one basis (i.e. the sum of the two weights is equal to one) between the auto-regressive part and the shock term in the GARCH specification. The distribution of weights on the two parts of the GARCH model in a certain period is dependent on their respective ability to predict a given benchmark volatility in the most recent period. For this reason we term our specification the benchmark volatility targeting-GARCH, or BVT-GARCH. Depending on the forecast performance of either the persistence term or the shock term, more weight is assigned to either persistence or shock term.

By doing so, we introduce a considerable amount of flexibility to the standard GARCH specification, while only introducing one additional parameter. The one-on-one characteristic of the weights is particularly attractive, because, besides from being simple, it embeds the typical characteristics of market volatility. It is well known that volatility is highly persistent, and thus that the driver of volatility is persistent. As such, if the auto-regressive (or shock) term of the GARCH specification does particularly well in describing the volatility benchmark in a certain period, it is also expected to do so in the next period. Therefore, the weight attached to this term is increased. Interestingly, our proposed setup introduces more volatility in volatility, which has proven to be of great importance in, e.g., the option pricing literature; see Christoffersen and Jacobs (2004) and Bams et al. (2009). Frijns et al. (2010a) use a setup comparable to ours to price options in a GARCH option pricing framework and show that the switching model is performing significantly better than the benchmark in both in-sample and out-of-sample.

We apply our model empirically to the 30 stocks in the Dow Jones Industrial Average for the period January 1996 to April 2007, where we use the realized variance (RV) as proposed by Andersen et al. (2003) as the benchmark volatility. Using our model, we find significant switching between the persistence and the shock term. Our model compares favourably to a standard GARCH model and can better capture the kurtosis and skewness observed in stock returns. Out-of-sample, we find that our model outperforms various alternative specifications, including a regime-switching GARCH and a GARCH with RV as an exogenous factor. In addition, we find that most of the improvement over the GARCH with RV is obtained for relatively volatile stocks.

The remainder of this paper is structured as follows. In Section 2, we describe our econometric model. Section 3 introduces the data and Section 4 presents our results. Finally, Section 5 concludes.

2. The Econometric Model

The motivation for the development of our model stems from the fact that standard GARCH models cannot deal with structural changes in the volatility process. Therefore, the apparent misspecification of the standard model is induced by a lack of flexibility. As a solution, several authors proposed to introduce time variation in the GARCH parameters (Lamoureux and Lastrapes, 1990). To deal with time variation in the parameters several authors have developed regime switching models (e.g. Cai, 1994; Hamilton and Susmel, 1994; Gray, 1996; and Klaassen, 2002). However, while these regime switching models generally perform better than standard GARCH models, the introduction of this time variation seems to be driven by empirical observation, rather than by economic rationale. In addition, regime switching GARCH models introduce a substantial number of extra parameters making estimation tedious.

The purpose of this section is to introduce a GARCH model that allows for time variation in parameters, but does so in a parsimonious and intuitive way. The specification of this GARCH model is based on the heterogeneous agents literature where agents (traders) maximize a certain objective function and switch between different trading rules to achieve this (see e.g. Franke, 2009). We apply this rationale to describe the volatility process in the following way. Each day risk managers try to make accurate forecasts of future volatility using a GARCH-type model. To measure their forecasting performance, they use a certain benchmark and evaluate their forecasting skills using their forecast error. Given their past forecasting performance, they may change the weights they give to certain forecasting rules. For example, in the GARCH framework, they may give more weight to the persistence term or to the shock term, depending on their ability to forecast volatility. Obviously, the forecasting performance of these rules does change over time, which translates into changing weights. It is this changing in the weights that we use to describe the volatility process and which introduces more or less volatility in the volatility process.

Specifically we consider the following GARCH-type model,[2]

$$
\begin{aligned}
r_t &= \mu + \sqrt{h_t}\,\varepsilon_t, \quad \varepsilon_t \sim N(0,1), \\
h_t &= \omega + w_t \beta h_{t-1} + (1 - w_t)\left(\alpha \left(\sqrt{h_{t-1}}\,\varepsilon_{t-1} \right)^2 \right),
\end{aligned} \tag{1}
$$

where r_t is the daily (log) return of a stock price and h_t describes the conditional variance of the return process. In this model, β is the persistence parameter and α measures the impact of news shocks. Risk managers, when making forecasts of volatility, use such a GARCH-type model, but can change the weights, w_t, they attach to either the persistence parameter or the shock parameter (w_t is based on time t-1 information). This w_t introduces time variation into the model, and it is easily seen that Equation (1) reduces to a standard GARCH model when w_t is constant.

The weights in Equation (1) given to the persistence parameter and the shock parameter depend on the relative forecasting performance of the persistence term and the shock term. We therefore define w_t as a multinomial switching process as originally introduced by Manski and McFadden (1981):[3]

$$
\begin{aligned}
w_t &= \frac{\exp\{\gamma \pi_{2t-1}\}}{\exp\{\gamma \pi_{1t-1}\} + \exp\{\gamma \pi_{2t-1}\}} = \left(1 + \exp\{\gamma \left(\pi_{2t-1} - \pi_{1t-1}\right)\}\right)^{-1}, \\
\pi_{1t-1} &= \left| \alpha \left(\sqrt{h_{t-2}}\,\varepsilon_{t-2} \right)^2 - RV_{t-1} \right|, \\
\pi_{2t-1} &= \left| \beta h_{t-2} - RV_{t-1} \right|.
\end{aligned} \tag{2}
$$

The weight, w_t, is bounded between zero and one and is a function of the performance of a forecasting rule that only considers shock terms, and a forecasting rule that is a function of the persistence term.

[2]We could easily extend the GARCH-specification in many directions. However, since our main focus is on the switching mechanism, we specify the simplest possible GARCH-specification.
[3]The unconditional volatility ω is excluded from the performance measures, π_{it}, as it is similar for both groups. Including ω would only lead to a level change in π_{1t} and π_{2t} but no change in the difference used in w_t.

To measure the relative performance of the forecasting rules, we need to introduce a benchmark. In this case we use the Realized Variance (RV) (see e.g. Andersen et al., 2003) as a benchmark to evaluate the forecasting performance.[4] Furthermore, we consider the absolute deviation from this benchmark as the main evaluation criterion.[5] The term, π_{1t-1}, considers the performance of the forecasting rule based on shock terms in forecasting RV, whereas the term π_{2t-1} considers the performance forecasting rule based on the persistence term in forecasting RV.

The γ-parameter, known in the heterogeneous agents literature as the intensity of choice parameter, measures the sensitivity of the switching process. The higher the absolute value of γ, the more weights will change in response to a difference between π_{1t-1} and π_{2t-1}. If $\gamma = 0$, w_t will be equal to ½ and constant through time, and Equation (1) reduces to a standard GARCH (1, 1). The other extreme, when $\gamma \to \pm\infty$, implies that weights are infinitely sensitive to the performance of the two forecasting rules and w_t will jump between zero and one. Generally, we expect $\gamma < 0$, which implies a positive feedback rule from the forecasting performance to the weights attached to each forecasting rule, i.e. if the volatility of the market is better forecasted using the rule based on the shock terms, then that rule receives more weight in the subsequent period and vice versa. As such, volatility clusters or stable periods can be reinforced not only by the persistence in the GARCH process but also by the increase or decrease in weight on the different elements of the model. Therefore, both upper and lower extremes in market volatility can be described more accurately compared with a normal GARCH. This causes the volatility of volatility temporarily to increase, which is especially important in describing volatility processes (see Christoffersen and Jacobs, 2004; and Bams et al., 2009).

The model proposed above is a specific version of a regime-switching GARCH model. Although it is not possible to rewrite Equations (1) and (2) into a specific and restricted version of a regime-switching GARCH (e.g. as a Generalized Regime Switching (GRS) model as proposed by Gray, 1996), we can show the analogy with a GRS-

[4]RV is often used as a benchmark volatility measure against which out-of-sample performance is compared (see e.g. Andersen et al., 2003; Frijns et al., 2010b, among others). However, the model is flexible enough to incorporate alternative benchmarks.

[5]Results are not qualitatively different when applying other functional forms such as squared deviations as the evaluation criterion.

GARCH. The GRS-GARCH where switching only occurs in the second moment and where the regimes follow a first-order Markov process is defined as follows,

$$\begin{aligned} h_t &= p_{1t}h_{1t} + (1 - p_{1t})h_{2t}, \\ p_{1t} &= (1 - Q)\left[\frac{g_{2t-1}(1 - p_{1t})}{g_{1t-1}p_{1t-1} + g_{2t-1}(1 - p_{1t})}\right] + P\left[\frac{g_{1t-1}p_{1t}}{g_{1t-1}p_{1t-1} + g_{2t-1}(1 - p_{1t})}\right], \\ g_{it-1} &= f(r_t \mid S_t = i). \end{aligned} \tag{3}$$

Where p_{1t} is the conditional probability of being in state 1, (1-Q) is the unconditional transition probability going from state 1 to state 2, P is the unconditional transition probability of staying in state 1, and g_{1t} and g_{2t} are the likelihoods of the return series being generated by regime 1 and 2, respectively. In this GRS-GARCH, the switching occurs between two different GARCH equations, h_{1t} and h_{2t}. In our GARCH specification the switching occur between a GARCH and an ARCH term. Furthermore, the switching in our model is a deterministic function of the forecasting performance of each rule, while the switching in Equation (3) is stochastic.

In the remainder of the paper, we will study the empirical merits of the BVT-GARCH model outlined above. We do this by contrasting its in-sample and out-of-sample performance of our model and compare its performance to several alternatives. The next section starts by describing the data and methodology.

3. Data and Methodology

The model outlined above is estimated and evaluated using daily data from the 30 individual stocks in the Dow Jones Industrial Average (DJIA). The sample covers January 1996 to April 2007, which corresponds to 2,820 trading days. Stock prices are taken from Datastream, where we obtain open and closing prices for the 30 stocks in the sample. Returns are calculated by $r_t = \log(p_{close}) - \log(p_{open})$. We step aside from the standard daily log-changes in closing prices because of the calculation of RV. The RV measure, which is based on intraday prices, only takes into account the volatility during trading hours and not the overnight period. Therefore, to match the return data with the RV data, one can either rescale the RV,[6] or focus on behaviour during trading hours only. The latter is chosen, as it does not make any assumptions on the relation between volatility during trading hours and non-trading hours. Table 1 provides descriptive statistics for return series of the stocks in the sample.

INSERT TABLE 1 HERE

The realized variances are computed following Andersen et al. (2003). To compute realized variances we obtain intraday data for the various stocks from the Taqtic database by SIRCA.[7] This database contains intraday data, which is provided by Reuters, for many stock markets around the world. Intraday data contain time of the trade recorded at one thousandth of a second, and the traded price (and many other details, which we do not exploit in this paper). Following Andersen et al. (2003) we compute the realized variance as the sum of squared intraday returns, where we aggregate returns to five-minute intervals, i.e.

$$RV_t = \sum_{j=1}^{J} r_{t,j}^2 , \tag{4}$$

where $r_{t,j}$ is the intraday return on day t for interval j. When forming the RV, there is a trade-off between accuracy (the intra-day squared returns converge to the actual volatility when the frequency goes to infinity, see Andersen et al., 2003), and microstructure noise at high frequencies (bid-ask bounce and other microstructure effects). We have chosen to use 5-minute returns, in line with Andersen et al. (2003) as a reasonable trade-off between microstructure noise and accuracy.[8] We report average annualized realized volatilities, i.e. $\sqrt{255 * RV_t}$, in the last column of Table 1.

The model outlined in Section 2 is estimated by quasi-maximum likelihood. As a result of the non-linear structure of the switching mechanism, estimation results of this type of setup have proven to be sensitive to starting values of the coefficients, as the optimization procedure tends to converge to a local optimum (see Gilli and Winker,

[6]Thus assuming a constant correlation between trading-hours volatility and overnight volatility.
[7]Securities Industry Research Centre of Asia-Pacific.
[8]Again, there are various alternative ways in which RV can be computed, with bias corrections and sampling at different frequencies. However, as this is not the main point of this paper, we resort to a simple approach that has been shown to produce reasonable estimates of volatility.

2004). To circumvent this problem, the model was first estimated with constant weights (i.e., $w_{i,t}=1/2$) to obtain starting values for the setup with switching mechanism. The starting values for γ were obtained by an extensive grid-search over the parameter space, using the log-likelihood

$$L(\theta \mid r_t) = -\frac{T}{2}\log(2\pi) - \frac{1}{2}\sum_{t=1}^{T}\log(h_t) - \frac{1}{2}\sum_{t=1}^{T}\left(\frac{(r_t-\mu)^2}{h_t}\right), \tag{5}$$

as criterion. In the end, results appeared to be relatively insensitive to changes in starting values. The performance of the BVT-GARCH is contrasted to that of the normal GARCH and the regime switching setup as outlined in Section 2, i.e. the GRS-GARCH. To control for the information advantage the BVT-GARCH has relative to the normal GARCH due to the use of an exogenous variable, the model is also compared to a setup in which RV is directly added in the variance equation. The resulting variance equation of the GARCH-RV model reads

$$h_t = \omega + \beta h_{t-1} + \alpha_1\left(\sqrt{h_{t-1}}\varepsilon_{t-1}\right)^2 + \alpha_2 RV_{t-1}, \tag{6}$$

while the BVT-GARCH-RV model is given by

$$h_t = \omega + w_t\beta h_{t-1} + (1-w_t)\left(\alpha_1\left(\sqrt{h_{t-1}}\varepsilon_{t-1}\right)^2 + \alpha_2 RV_{t-1}\right), \tag{7}$$

where

$$\begin{aligned} w_t &= \frac{\exp\{\gamma\pi_{2t-1}\}}{\exp\{\gamma\pi_{1t-1}\}+\exp\{\gamma\pi_{2t-1}\}} = \left(1+\exp\{\gamma(\pi_{2t-1}-\pi_{1t-1})\}\right)^{-1}, \\ \pi_{1t-1} &= \left|\alpha_1\left(\sqrt{h_{t-2}}\varepsilon_{t-2}\right)^2 + \alpha_2 RV_{t-2} - RV_{t-1}\right|, \\ \pi_{2t-1} &= \left|\beta h_{t-2} - RV_{t-1}\right|. \end{aligned} \tag{8}$$

We have added RV to the shock term because the elements are conceptually equivalent, i.e., both are exogenous to the volatility process given by h_t.[9]

[9] We did not add a third group consisting of the RV-term alone for comparability to the benchmark model (2).

4. Results

The results section is split up into three parts. First, the (in-sample) estimation results are presented, including a comparison of our model with a number of alternative setups. Second, to gain some more insights into the model, a close-up of one of the stocks, IBM, is presented. The third sub-section looks into the out-of-sample results.

4.1 In-sample Results

Table 2 presents the estimation results of empirical model formed by Equations (1) and (2) estimated for the thirty DJIA stocks.

INSERT TABLE 2 HERE

The estimated coefficients ω, α, and β are basically comparable to the standard GARCH coefficients as known from the literature, but note that α and β should be divided by two to make them directly comparable to standard coefficients, because $w_{i,t}$ is, approximately, centred around one half. The intercept, ω, is usually small and not significantly different from zero. This is different from the standard GARCH model, where ω is usually found to be small, but positive and significant. Apparently, the switching mechanism lowers the unconditional volatility. This is a result of the fact that the model is locally unstable but globally stable. The IGARCH model has no intercept, but is still globally stable; see Nelson (1990). As such, our model represents a combination between the regular GARCH and IGARCH. The α-coefficient is in the proximity of 0.10 and highly significant; the β-coefficient is close to two and highly significant as well.

Concerning the switching function, we find that the sensitivity parameter γ is negative for all stocks. This implies that there is indeed a positive feedback rule, i.e., the element of the GARCH equation that is closest to the benchmark volatility in period t, will receive more weight in period t+1. Furthermore, γ is highly significant for all stocks, which serves as a first indication of the fact that adding this flexibility to the GARCH is beneficial to the in-sample fit.

INSERT TABLE 3 HERE

The likelihoods of a number of relevant alternative GARCH specifications are given in Table 3.[10] First of all, the BVT-GARCH outperforms the standard GARCH (1, 1) in terms of the average. This result holds for all thirty stocks. However, the average log-likelihood of the GRS-GARCH is higher than that of the BVT-GARCH, but this is expected as the GRS-GARCH has considerably more parameters. The GARCH-RV shows a better in-sample fit than the BVT-GARCH in terms of the log-likelihood. However, after introducing time-varying weights to this setup, as in Equation (6), the BVT-GARCH-RV is again outperforming. Generally, we can conclude that the BVT-GARCH is performing well compared with the alternative specifications.

Comparison on the basis of likelihood, however, may not provide a clear picture of how well these models perform relative to each other as likelihood increases when more parameters are added. To compare models more equally, we compute the Akaike and Schwartz Bayesian Information criteria for each model and report the number of "winners" (i.e. the times the model is best under the specific criterion) in Table 3. According to the AIC, the full regime switching GARCH (GRS-GARCH) wins most of the time.[11] Second to that are the BVT-GARCH models, the BVT-GARCH-RV winning 8 times and the BVT-GARCH winning twice. When turning to the Schwartz Bayesian criterion, which penalizes more for degrees of freedom, we observe that the GARCH-RV performs best winning 13 times. However, our BVT-GARCH models still stand up well, the BVT-GARCH winning 5 times and the BVT-GARCH-RV winning 3 times.

INSERT TABLE 4 HERE

Table 4 complements Table 3 by providing descriptive statistics for the normalized returns, ε_t. For all statistics given in the Table, the BVT-GARCH brings the statistics closer to normality compared to the GARCH (1, 1).[12] The maximum and minimum are less extreme for the BVT-GARCH. Both the skewness is closer to zero and the kurtosis closer to three, such that the JB-statistic is considerably lower for the BVT-GARCH.

[10]Likelihood ratio tests are possible when models are embedded. All models reduce to the GARCH (1, 1) benchmark under parameter restrictions.

[11]It is a well-known result that more sophisticated, but misspecified models achieve good in-sample results by overfitting the data, but have less predictive power in out-of-sample comparison.

[12]In fact, results are somewhat distorted due to a small number of extreme outliers. The descriptive statistics of normalized returns for some stocks are extremely affected by individual observations. Without these observations, the BVT-GARCH comes out stronger.

The 5% tail index is also larger for the BVT-GARCH. A similar conclusion holds for the autocorrelation pattern in the squared normalized returns, as presented in Panel B; probabilities for the BVT-GARCH are consistently and considerably higher.

4.2 Example: IBM

INSERT FIGURE 1 HERE

Figure 1 illustrates the behaviour of our model for the case of one typical stock, IBM. Plot A displays the effect that introducing the switching process has on the volatility process by illustrating the volatility process induced by the BVT-GARCH (lower line) and the difference with the GARCH (1, 1) (upper line). Clearly, the introduction of time variation gives much more flexibility to the volatility process. Comparing the orders of magnitude, it becomes clear that the maxima (minima) are up to 50% higher (lower) for the BVT-GARCH relative to the GARCH (1, 1). This comes forward especially in the second half of 2002, where the volatility from the BVT-GARCH is much higher than of the GARCH (1, 1). Differences between the two volatility processes are especially pronounced during periods of large changes in volatility.

Plot B displays the weights w_t and the volatility h_t over time. Changes in weights coincide with changes in volatility. This can be explained by the fact that as volatility itself is volatile, the element that describes the benchmark volatility best is also more likely to change through time. In the final part of the sample, where volatility itself becomes more stable, the weights also stabilize around one half.

Plot C illustrates in more detail where the differences between the BVT-GARCH and the GARCH (1, 1) exactly take place. The scatter of the two conditional volatilities clearly diverges as the volatility increases. That is, the differences between the two models are most pronounced when volatility is high, which coincides with periods in which volatility in volatility is high.

Plot D illustrates the relation between the relative performance of the forecasting rules and the resulting weights. Clearly, there is an upward slope indicating that the better performing rule attracts more weight. Furthermore, the slight S – shape, which is induced by the exponential function, can be seen.

Plot E, finally, shows the histogram of the weights for IBM. On average, *w* is just above one half, indicating that the persistence element of the GARCH is slightly dominating. In addition, it can be seen that the weight is typically around one half, with large spikes towards zero and one, as could already be seen in Plot B. The autocorrelation (AC) in the weights is equal to 0.38, indicating the persistence not only in the volatility but also in the distribution of weights over the persistence and shock terms.

4.3 Out-of-Sample Results

In addition to the in-sample performance of our proposed model, the out-of-sample properties are considered. This is done by applying the following methodology. First, the models are estimated for the first 1,400 days, i.e., half of the sample, and subsequently one-period forecasts are made for the remaining 1,420 days. The stocks Verizon (VZ) and Exxon Mobile (XOM) are not included in the analysis as both were introduced in the DJIA after 1996.[13]. Table 5 reports the mean forecast error (ME); mean squared error (MSE); root mean squared error (RMSE); mean absolute error (MAE); mean absolute percentage error (MAPE); heteroskedasticity adjusted mean squared error (HMSE); the R^2 of the regression given by $\log RV_t = \alpha + \beta \log E_{t-1}(h_t) + \varepsilon_t$; and the number of cases in which the BVT-GARCH beats the alternative.

INSERT TABLE 5 HERE

Generally, for all forecasting performance measures, the ranking is the same. The BVT-GARCH comes out first, followed by the GARCH-RV, the GRS-GARCH, and the GARCH (1, 1). The one exception to this pattern is given in Panel F, the HMSE. Here, the BVT-GARCH performs worst. Note, however, that a large part of this result is driven by an outlier; the maximum value of the BVT-GARCH is more than twice as large as the second worst model, the GARCH-RV. When considering the last row of each panel, we observe that the BVT-GARCH beats the GARCH and GRS-GARCH

[13] Note that in total three stocks were excluded in the forecasting exercise due to lack of data.

consistently on all performance measures and for most performance measures we beat the GARCH-RV.

INSERT TABLE 6 HERE

Finally, Table 6 provides evidence of the fact that the BVT-GARCH model is especially useful in high volatile regimes. In times of high volatility, and high volatility of volatility, one would expect that the changing weights provide their biggest advantage as it is especially in risky situations where the extra flexibility is granted. The reinforcement of the persistence in variance by means of changing weights is particularly beneficial as there is ample change in the behaviour of volatility. We address this issue by examining the cross-sectional relationship between forecasting performance and volatility. Specifically, we compare the forecasting performance of the BVT-GARCH model versus the second best, the GARCH-RV model cross-sectionally. That is, we compare the forecasting performance of the BVT-GARCH versus the GARCH-RV across stocks and link this difference to the volatility of the stock.

Table 6 presents the cross-sectional correlations between the differences in forecasting performance of the BVT-GARCH model versus the second best, the GARCH-RV model. One can observe a clear pattern in the correlations. For the forecast error measures, apart from the ME, there is a strong negative correlation. That is, the higher the (unconditional) volatility in the stock, the better the forecasting performance of the BVT-GARCH vis-à-vis the GARCH-RV. The correlation is positive for the R^2, as the fit is a positive measure of forecasting ability while the forecasting error is a negative measure. The exception to this rule is the mean error ME. This can be explained by the fact that typical patterns tend to be eliminated as positive and negative values are levelled out.

5. Conclusion

The Nobel-price winning GARCH methodology has seen a large number of extensions. Of particular interest is the introduction of time-varying coefficients, which serves as a methodology of curving the (excessive) persistence of volatility in financial market returns. Typically, this time variation is introduced by means of purely stochastic processes, such as a regime switching approach. Alternatively, GARCH models are introduced that mix a fast with a slow moving process.

In this paper, we introduce a parsimonious methodology of time variation in the coefficients to introduce more flexibility and dynamics into the conditional volatility model. We propose a methodology that enforces the typical characteristics of a standard GARCH. To be more specific, we introduce time-varying weights on the persistence and shock coefficients. The weights move according to a multinomial logit process that gives more importance to the factor, i.e. persistence or shock, which better predicted benchmark volatility in the previous period. As such, we refer to the benchmark volatility targeting or BVT-GARCH model.

Estimating the model on the thirty individual stocks in the Dow Jones Industrial Average, it is shown that the BVT-GARCH model outperforms alternative setups. Using only one additional parameter, the model proves to be better in both in-sample and out-of-sample tests. In-sample, we find increased likelihood values, and normalized returns that are considerably closer to be being Gaussian relative to normalized returns from alternative models. Out-of-sample, using a broad range of forecasting performance measures, the BVT-GARCH comfortably outperforms a number of alternative models, including a regime switching GARCH and controlling for potential information advantages. Furthermore, the model proves to be particularly useful in times of high volatility, or high volatility in volatility.

References

Amado, C. and T. Teräsvirta (2008). Modelling Conditional and Unconditional Heteroskedasticity with Smoothly Time-Varying Structure. Working paper.

Andersen, T. G., T. Bollerslev, F.X. Diebold, and P. Labys (2003). Modelling and Forecasting Realized Volatility, *Econometrica* 71: 529-626.

Bams, D., T. Lehnert and C. Wolff (2009). Loss Functions in Option Valuation: A Framework for Selection. *Management Science* 55: 853-862.

Bollerslev, T. (1986). Generalized Autoregressive Conditional Heteroskedasticity, *Journal of Econometrics* 31: 307 – 327.

Brock, W. and C. Hommes (1997). A Rational Route to Randomness, *Econometrica* 69: 1059-1095.

Brock W. and C. Hommes (1998). Heterogeneous Beliefs and Routes to Chaos in a Simple Asset Pricing Model, *Journal of Economic Dynamics and Control* 22: 1235-1274.

Cai, J. (1994). A Markov Model of Switching-Regime ARCH, *Journal of Business and Economics Statistics* 12: 309 – 316.

Christoffersen, P., Jacobs, K. (2004). Which GARCH Model for Option Valuation?, *Management Science* 50: 1204-1221.

Dahlhaus R. and S. Subba Rao (2006). Statistical Inference for Time-Varying ARCH Processes, *Annals of Statistics*, 34, 1075-1114.

Diebold, F.X. (1986). Modelling the Persistence of Conditional Variances: A comment, *Econometric Reviews* 5: 51-56.

Van Dijk, D., T. Teräsvirta and P.H. Franses (2002). Smooth transition autoregressive models - a survey of recent developments, *Econometric Reviews* 21: 1-47.

Engle, R.F. (1982). Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of U.K. Inflation, *Econometrica* 50: 987 – 1008.

Engle, R.F. and J.G. Rangel (2008). The Spline-GARCH Model for Low-Frequency Volatility and its Global Macroeconomics Causes, *Review of Financial Studies* 21: 1187 – 1222.

Franke, R. (2009). Applying the method of simulated moments to estimate a small agent-based asset pricing model, *Journal of Empirical Finance* 16: 804-815.

Frijns, B., Lehnert, T., and Zwinkels, R.C.J. (2010a). Behavioral Heterogeneity in the Option Market, *Journal of Economic Dynamics and Control* 34: 2273-2287.

Frijns, B. Tallau, C., and Tourani-Rad, A. (2010b). The Information Content of Implied Volatility: Evidence from Australia, *Journal of Futures Markets* 30: 134-155.

Gilli, M. and P. Winker (2004). A Global Optimization Heuristic for Estimating Agent Based Models, *Computational Statistics and Data Analysis* 42: 299 – 312.

Gray, Stephen F. (1996), Modelling the Conditional Distribution of Interest Rates as a Regime-Switching Process, *Journal of Financial Economics* 42: 27-62.

Hamilton, J. and R. Susmel (1994). Autoregressive Conditional Heteroskedasticity and Changes in Regime, *Journal of Econometrics* 2: 493 – 530.

Klaassen, F. (2002). Improving GARCH Volatility Forecasts with Regime-Switching GARCH, *Empirical Economics* 27: 363-394.

Lamoureux, C. and W. Lastrapes (1990). Heteroskedasticity in Stock Return Data: Volume Versus GARCH Effects, *Journal of Finance* 45: 221 – 229.

Manski, C.F. and D. McFadden (1981). Alternative Estimators and Sample Designs for Discrete Choice Analysis, in: C.F. Manski and D. McFadden (eds.), *Structural Analysis of Discrete Data and Econometric Applications*, MIT Press: Cambridge.

Nelson, D.B. (1990). Stationarity and Persistence in the GARCH(1,1) Model, *Econometric Theory* 6: 318-334.

Schwert, G. (1989). Why Does Stock Market Volatility Change over Time? *Journal of Finance* 44: 1115 – 1153.

Tables & Figures

Table 1: Descriptive Statistics Returns

	Mean	Max	Min	St.Dev	Skew	Kurt	Q-ac	Q-ac^2	RV
AA	0.034	8.338	-13.286	1.908	-0.421	5.566	3.731	21.14	0.273
AIG	0.005	10.436	-10.713	1.619	-0.092	6.739	3.688	91.83	0.224
AXP	-0.018	12.063	-9.883	1.828	0.013	6.103	0.339	79.62	0.257
BA	-0.022	10.739	-10.276	1.73	-0.009	6.095	1.362	150.12	0.256
CAT	0.021	7.85	-9.97	1.799	-0.11	5.05	0.089	72.9	0.26
C	0.031	13.005	-15.861	1.824	-0.281	8.605	9.297	320.94	0.263
DD	-0.019	9.465	-10.561	1.658	-0.275	6.155	0.002	80.82	0.249
DIS	-0.018	11.804	-9.304	1.75	-0.045	5.69	1.685	70.28	0.27
GE	-0.018	8.38	-11.123	1.59	-0.257	6.769	0.07	101.43	0.228
GME	0.122	8.279	-7.974	1.86	-0.131	4.561	0.032	93.57	0.253
HD	0.008	11.303	-12.128	1.861	-0.176	5.985	0.879	89.04	0.27
HON	0.012	19.839	-14.91	1.957	0.209	11.383	0.002	94.89	0.269
HPQ	-0.075	12.995	-15.887	2.3	-0.185	6.19	1.159	12.09	0.321
IBM	-0.034	10.37	-12.209	1.702	-0.107	7.103	9.242	58.26	0.231
INTC	0.051	12.063	-20.493	2.36	-0.279	6.857	20.57	69.46	0.333
JNJ	-0.028	7.411	-7.711	1.313	-0.181	5.301	0.855	113.37	0.202
JPM	-0.034	16.364	-24.429	1.97	-0.949	16.706	6.586	268.62	0.254
KO	-0.074	11.158	-9.175	1.41	-0.104	7.378	0.342	41.73	0.215
MCD	-0.077	11.431	-14.284	1.621	-0.108	9.042	1.624	46.22	0.25
MMM	-0.013	7.045	-8.38	1.398	-0.131	5.83	0.022	33.11	0.213
MO	0.013	14.714	-12.027	1.772	0.503	11.671	1.362	88.31	0.241
MRK	-0.053	11.166	-9.347	1.591	0.142	6.879	0.525	39.37	0.227
MSFT	-0.022	8.907	-11.011	1.863	-0.199	5.272	10.16	164.51	0.266
PFE	0.01	8.829	-10.341	1.735	-0.126	5.277	0.957	43.97	0.246
PG	-0.13	9.237	-10.516	1.416	0.011	8.353	23.18	235.12	0.212
T	-0.015	6.881	-10.752	1.74	-0.213	5.24	0.41	67.23	0.265
UTX	0.022	16.462	-7.339	1.621	0.468	8.918	0.356	19.29	0.237
VZ	-0.021	6.312	-8.545	1.571	-0.494	5.801	0.335	65.8	0.227
WMT	0.017	8.719	-12.915	1.743	-0.226	6.444	3.051	53.63	0.262
XOM	-0.022	6.8	-11.605	1.359	-0.34	7.8	2.536	12.28	0.202

Notes: Table presents the descriptive statistics of the raw returns. Q-ac represents the Ljung-Box Q-statistic for autocorrelation in the returns; Q-ac^2 is the Q-statistic for autocorrelation in the squared returns. RV represents the annualized average realized volatility.

Table 2: BVT-GARCH In-Sample Estimation Results

	ω	α	β	γ	LogL
AA	0.000	0.038[a]	1.977[a]	-0.540[a]	7371.21
	(0.000)	(2.707)	(103.39)	(-5.298)	
AIG	0.000	0.118[a]	1.919[a]	-1.582[a]	7987.07
	(0.158)	(6.699)	(103.27)	(-5.657)	
AXP	0.000	0.116[a]	1.913[a]	-1.760[a]	7736.27
	(0.546)	(5.844)	(87.06)	(-5.668)	
BA	0.000	0.125[a]	1.928[a]	-2.908[a]	7684.10
	(0.165)	(5.788)	(62.39)	(-8.707)	
CAT	0.000	0.077[a]	1.994[a]	-3.075[a]	7524.73
	(0.000)	(3.750)	(40.50)	(-7.069)	
CG	0.002[a]	0.226[a]	1.796[a]	-1.897[a]	7732.54
	(2.601)	(10.98)	(77.36)	(-7.046)	
DD	0.000	0.028	1.975[a]	-3.009[a]	7874.95
	(0.318)	(1.605)	(72.30)	(-7.089)	
DIS	0.005[a]	0.098[a]	1.848[a]	-2.400[a]	7708.95
	(3.001)	(6.547)	(62.12)	(-9.682)	
GE	0.000	0.099[a]	1.925[a]	-2.380[a]	8104.97
	(0.000)	(6.270)	(101.90)	(-8.276)	
GME	0.000	0.112[a]	1.978[a]	-2.007[a]	7408.33
	(0.000)	(5.218)	(40.93)	(-6.717)	
HD	0.000	-0.001	2.017[a]	-1.757[a]	7564.62
	(0.000)	(-0.066)	(99.66)	(-14.57)	
HON	0.000	0.208[a]	1.858[a]	-1.505[a]	7479.15
	(0.000)	(15.89)	(86.39)	(-11.93)	
HPQ	0.000	0.066[a]	2.012[a]	-1.989[a]	6910.04
	(0.000)	(3.156)	(48.57)	(-12.24)	
IBM	0.000	0.114[a]	1.933[a]	-2.076[a]	7905.70
	(0.000)	(7.145)	(84.11)	(-6.037)	
INTC	0.000	0.098[a]	1.961[a]	-1.293[a]	6892.63
	(0.000)	(4.628)	(54.41)	(-9.294)	
JNJ	0.001[a]	0.122[a]	1.866[a]	-4.031[a]	8526.75
	(2.883)	(5.596)	(76.90)	(-10.69)	
JPM	0.000	0.143[a]	1.903[a]	-1.597[a]	7660.35
	(0.000)	(9.618)	(99.00)	(-7.814)	
KO	0.001	0.113[a]	1.885[a]	-5.896[a]	8421.69
	(1.304)	(5.551)	(73.47)	(-11.02)	
MCD	0.001	0.093[a]	1.939[a]	-4.488[a]	7831.25
	(0.515)	(6.175)	(71.58)	(-11.42)	
MMM	0.000	0.050[a]	1.982[a]	-4.187[a]	8268.82
	(0.000)	(2.550)	(60.69)	(-9.140)	
MO	0.000	0.141[a]	1.956[a]	-2.376[a]	7701.09
	(0.000)	(8.023)	(78.43)	(-11.00)	
MRK	0.000	0.050[a]	2.072[a]	-3.865[a]	7798.36
	(0.070)	(2.712)	(43.69)	(-10.35)	
MSFT	0.000	0.152[a]	1.923[a]	-3.141[a]	7654.91
	(0.000)	(5.542)	(57.76)	(-9.380)	
PFE	0.000	0.119[a]	1.964[a]	-3.183[a]	7676.86
	(0.000)	(5.456)	(46.48)	(-7.911)	
PG	0.002[a]	0.051[a]	1.893[a]	-5.230[a]	8416.41
	(3.185)	(4.353)	(63.65)	(14.97)	
T	0.000	0.065[a]	1.936[a]	-0.148[a]	7753.56
	(0.731)	(6.910)	(210.44)	(-3.100)	
UTX	0.000	0.089[a]	1.949[a]	-2.421[a]	7965.70
	(0.000)	(8.064)	(82.01)	(-7.352)	
VZ	0.001	0.032	1.955[a]	-3.561[a]	4963.98
	(0.641)	(1.324)	(55.29)	(-9.049)	
WMT	0.000	0.038	1.993[a]	-3.237[a]	7783.82
	(0.000)	(1.427)	(62.47)	(-13.748)	
XOM	0.002	0.061[a]	1.919[a]	-2.678[a]	5503.73
	(1.239)	(2.774)	(54.77)	(-5.186)	

Notes: Table presents the estimation results of the model given by Equations (1) and (2). T-values are reported in parentheses and [a] indicates significance at the 1%.

Table 3: In-sample fit of BVT-GARCH compared with other specifications

	# Coeff	Average(LogL)	Min(LogL)	Max(LogL)	AIC Best	BIC Best
BVT-GARCH	5	7593.67	4963.98	8526.75	2	5
GARCH(1,1)	4	7555.98	4946.37	8485.27	0	0
GRS-GARCH	9	7618.98	4958.26	8537.55	15	9
GARCH-RV	5	7598.50	4973.44	8532.32	5	13
BVT-GARCH-RV	6	7602.75	4973.52	8533.89	8	3

Notes: Table presents the average, minimum, and maximum log-likelihood of the in-sample estimation of the different setups over the 30 individual stocks.

Table 4: Properties of the BVT-GARCH

Panel A: Descriptive statistics normalized returns (averages)

		Raw Returns	GARCH(1,1)	BVT-GARCH
Maximum		6.7404	5.4955	5.2776
Minimum		-6.1544	-6.0603	-5.8186
Skewness		0.1364	-0.0166	-0.0157
Kurtosis		7.1584	5.3660	5.1057
Jarque-Bera		2732.37	1201.06	1078.27
L Tail Index	5%	3.148	3.613	3.768
R Tail Index	5%	3.087	3.671	3.757

Panel B: Auto-correlation squared normalized returns (averages)

	Raw Returns		GARCH(1,1)		BVT-GARCH	
lag	AC	Prob	AC	Prob	AC	Prob
1	0.1695	0.0000	0.0401	0.1673	0.0246	0.2982
2	0.1217	0.0000	0.0144	0.2004	-0.0003	0.3920
3	0.1116	0.0000	-0.0008	0.2337	-0.0080	0.4209
4	0.0910	0.0000	-0.0041	0.2663	-0.0062	0.4468
5	0.0897	0.0000	-0.0053	0.2965	-0.0083	0.4709

Notes: Table presents descriptive statistics of the normalized returns, i.e. ε_t, averaged over the thirty stocks. The tail index is estimated by the Hill estimator. AC denotes auto-correlation.

Table 5: Forecasting Performance of Various Specifications

Panel A: ME (x10^5)				
	BVT-GARCH	GARCH	GARCH-RV	GRS-GARCH
Mean	3.007	6.052	3.920	5.220
Median	2.615	4.820	3.150	3.837
Min	-10.100	-1.610	-2.400	-3.607
Max	10.100	16.600	11.700	17.871
BVT-GARCH wins		24	16	20
Panel B: MSE (x10^8)				
	BVT-GARCH	GARCH	GARCH-RV	GRS-GARCH
Mean	7.606	9.341	7.964	8.415
Median	4.680	6.430	4.900	5.546
Min	0.563	1.020	0.618	1.696
Max	25.800	31.300	29.900	29.675
BVT-GARCH wins		25	17	23
Panel C: RMSE (x10^3)				
	BVT-GARCH	GARCH	GARCH-RV	GRS-GARCH
Mean	0.253	0.284	0.257	0.269
Median	0.217	0.254	0.222	0.235
Min	0.075	0.101	0.079	0.130
Max	0.508	0.559	0.547	0.545
BVT-GARCH wins		25	17	23
Panel D: MAE (x10^3)				
	BVT-GARCH	GARCH	GARCH-RV	GRS-GARCH
Mean	0.121	0.151	0.123	0.140
Median	0.113	0.145	0.111	0.131
Min	0.046	0.079	0.052	0.080
Max	0.234	0.259	0.232	0.295
BVT-GARCH wins		26	16	22
Panel E: MAPE				
	BVT-GARCH	GARCH	GARCH-RV	GRS-GARCH
Mean	0.999	1.383	1.006	1.272
Median	0.815	1.284	0.869	0.929
Min	0.500	0.581	0.412	0.581
Max	2.365	3.404	2.471	2.885
BVT-GARCH wins		25	14	21
Panel F: HMSE				
	BVT-GARCH	GARCH	GARCH-RV	GRS-GARCH
Mean	0.806	0.636	0.599	0.478
Median	0.438	0.472	0.407	0.410
Min	0.240	0.348	0.231	0.261
Max	5.249	1.561	2.582	1.268
BVT-GARCH wins		21	7	17
Panel G: R^2				
	BVT-GARCH	GARCH	GARCH-RV	GRS-GARCH
Mean	0.584	0.479	0.579	0.534
Median	0.588	0.504	0.599	0.560
Min	0.367	0.217	0.334	0.241
Max	0.742	0.693	0.764	0.741
BVT-GARCH wins		27	11	23

Notes: Table presents the forecasting performance of different model setups over the 27 individual stocks measured by the mean forecast error (ME); mean squared error (MSE); root mean squared error (RMSE); mean absolute error (MAE); mean absolute percentage error (MAPE); heteroskedasticity adjusted mean squared error (HMSE); and the R^2 of the regression given by $\log RV_t = \alpha + \beta \log E_{t-1}(h_t) + \varepsilon_t$.

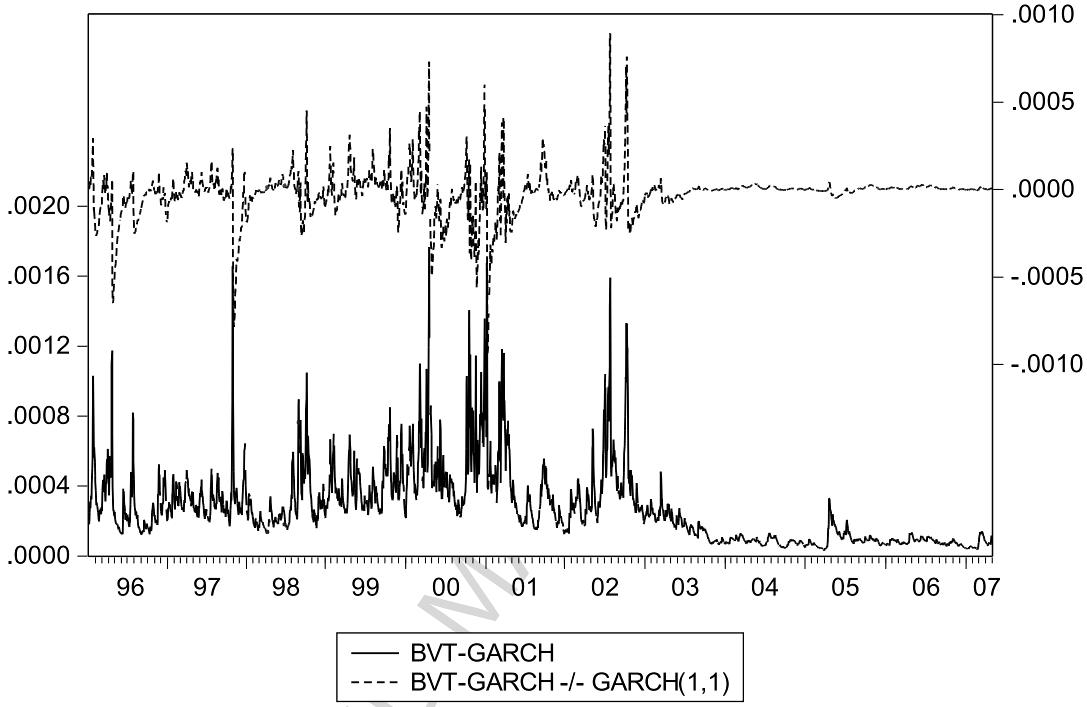
Table 6: Differences in Forecast Errors and Volatility

	R2	ME	MSE	RMSE	MAE	MAPE	HMSE
Correlation	0.1580	0.1833	-0.3881	-0.3733	-0.4373	-0.4317	-0.3465

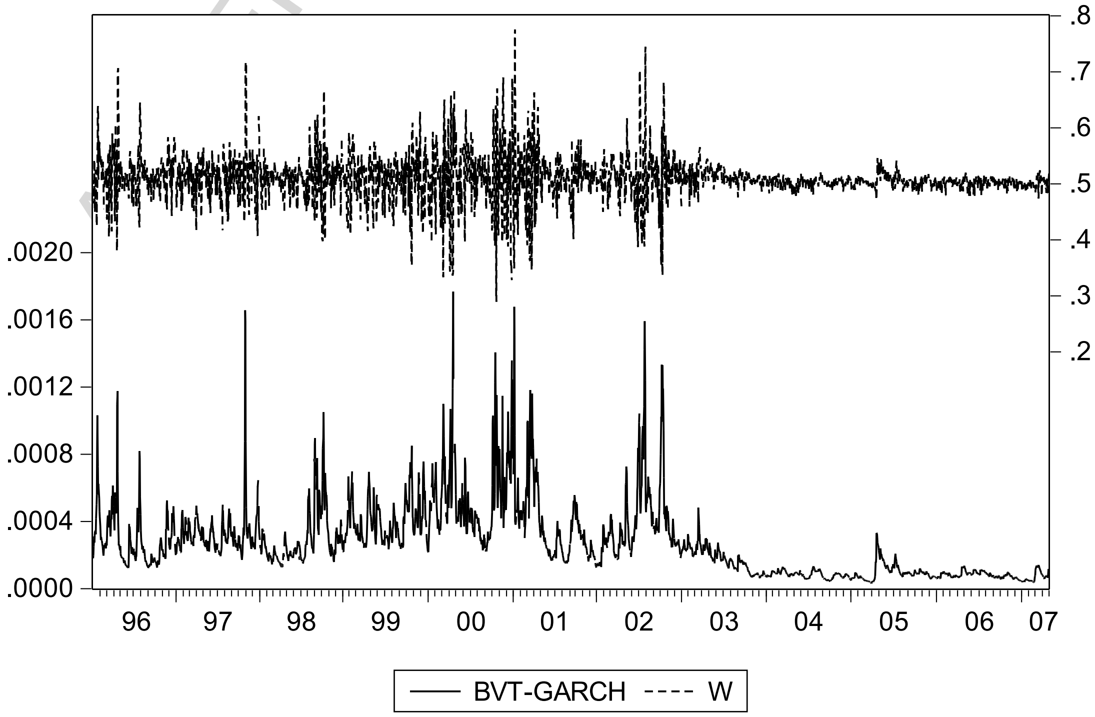
Notes: This table report the correlations between the difference between the forecast error of the BVT-GARCH and the GARCH-RV, and stock price volatility.

Figure 1: Properties of the BVT-GARCH model: IBM

Plot A: Conditional Volatility Switching minus Static GARCH

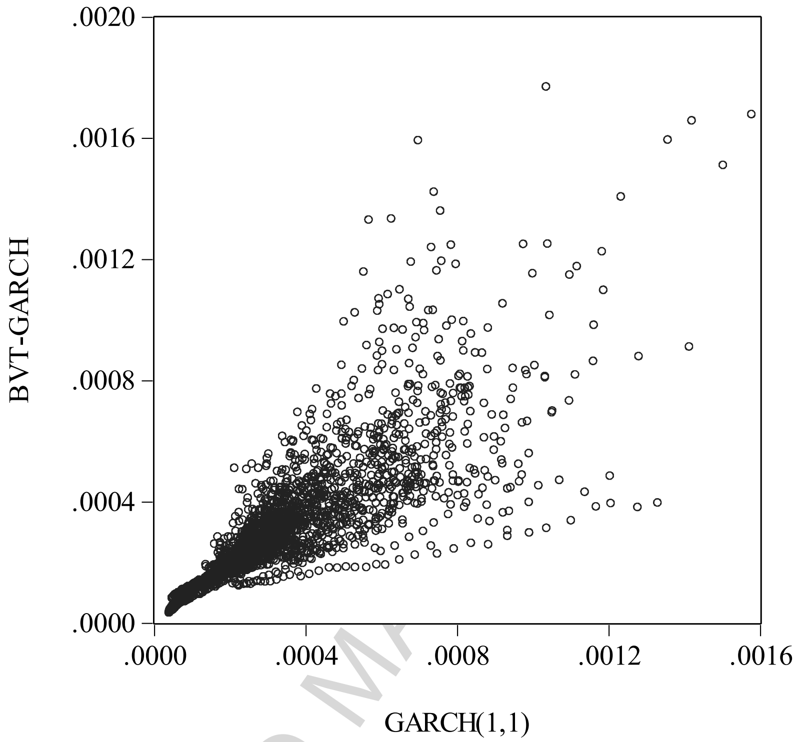


Plot B: Weights and Conditional Volatility

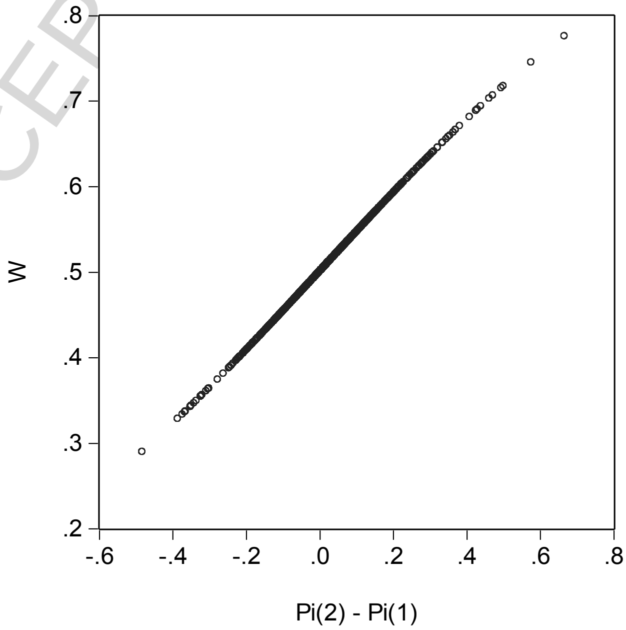


.

Plot C: Conditional Volatility Static versus BVT-GARCH



Plot D: Weights and Relative Forecast Accuracy



Plot E: Descriptive Statistics of the Weights